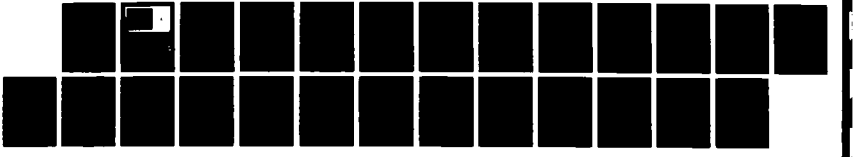
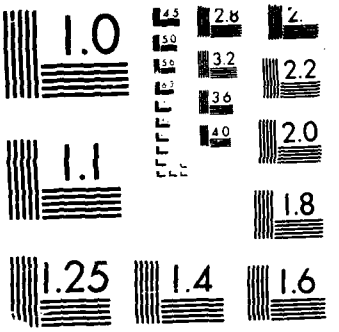
AD-A193 253 A SIMPLE EXPLANATION OF THE STOKES PHENOMENON(U) 1/1
WISCONSIN UNIV-MADISON CENTER FOR MATHEMATICAL SCIENCES
R E MEYER NOV 87 CMS-TSR-88-18 AFOSR-87-0194

UNCLASSIFIED F/G 12/1 NL





CMS Technical Summary Report #88-18

# A SIMPLE EXPLANATION OF THE STOKES PHENOMENON

Richard E. Meyer



UNIVERSITY OF WISCONSIN

CENTER FOR THE MATHEMATICAL SCIENCES

**Center for the Mathematical Sciences**
**University of Wisconsin—Madison**
**610 Walnut Street**
**Madison, Wisconsin 53705**

November 1987

(Received October 16, 1987)



**Approved for public release**
**Distribution unlimited**

Sponsored by

National Science Foundation
1800 G Street, N.W.
Washington, DC 20550

Air Force Office of
Scientific Research
Bolling AFB
Washington, DC

UNIVERSITY OF WISCONSIN-MADISON
CENTER FOR THE MATHEMATICAL SCIENCES

# A SIMPLE EXPLANATION OF THE STOKES PHENOMENON

R.E. Meyer*

CMS Technical Summary Report #88-18
November 1987

## ABSTRACT

The Stokes Phenomenon is known to be a pervasive feature of asymptotics, but its explanation in the literature is obscured by intricate and lengthy technicalities. This article presents a simpler approach to its understanding and treatment as a natural aspect of a well-motivated characterization of functions by approximants of different multivaluedness.

AMS (MOS) Subject Classifications: 34E20, 41A60, 81F05

Key words: Asymptotics, Waves.



*Mathematics Department, University of Wisconsin-Madison, Madison, WI 53706.

This work was supported in part by the National Science Foundation under Grant No. DMS-8521687 and by the U.S. Air Force Office of Research under Grant No. AFOSR-87-0194.

# A SIMPLE EXPLANATION OF THE STOKES PHENOMENON

Richard E. Meyer

## 1. Introduction

The Stokes Phenomenon plays a pervasive role in special functions and mathematical asymptotics, but access to its understanding from the literature is rather difficult because it emerges there as a late, very technical byproduct of relatively intricate analysis. In Olver's monograph [16], for instance, it is first mentioned on p. 241, but only with a reference to p. 481 for its explanation. The great merit of that explanation is obscured, moreover, by its limitation to very restricted circumstances [16]. The Author has never seen a simple explanation of the Phenomenon in its general context. The pervasiveness of the phenomenon, on the other hand, indicates that it should be generic and possess fundamental roots. It may therefore be useful to offer here a direct explanation of the phenomenon as a natural consequence of a decision to characterize functions by the help of approximating functions of a multivaluedness different from that of the functions to be characterized.

The next, brief Section gives a very simple example illustrating the root of the Stokes Phenomenon so clearly that many Readers may feel no need to proceed further. Section 3 tries to explain the many strong reasons for the choice of a characterization of functions by approximants of different multivaluedness. It also introduces a framework favorable for a fruitful discussion of the main qualitative and quantitative aspects of the

---

*Mathematics Department, University of Wisconsin-Madison, Madison, WI 53706.

This work was supported in part by the National Science Foundation under Grant No. DMS-8521687 and by the U.S. Air Force Office of Research under Grant No. AFOSR-87-0194.

Phenomenon. Section 4 discusses the location of the Phenomenon and adds some cautions for the interpretation of its most common statement in the mathematical literature. The last Section sketches a direct approach from the root of the Phenomenon to its quantitative prediction.

## 2. Multivalued Approximation

It is best to start with the simplest, concrete example and proceed then to explain why it illustrates matters of great generality. Airy's equation is

$$d^2y/dz^2 = z\, y(z) \,. \tag{1}$$

The solutions $y(z)$ of (1) are approximated for large $|z|$ by linear combinations of

$$u_{\pm} = z^{-1/4} \exp(\pm x), \qquad x = 2z^{3/2}/3 \,. \tag{2}$$

The meaning of approximation is here rigorous [16] and turns out to be quantitively useful even when $\epsilon$ is only moderately small.

Obviously, $u_+$ and $u_-$ are multivalued functions of the complex variable $z$ with a branch point at $z = 0$. By contrast, the solutions $y(z)$ of (1) are <u>entire</u> functions of $z$ because the coefficient function in (1) is $z$, which is entire. Therefore, as we go once around the point $z = 0$, $y(z)$ will return to its original value, but $u_+$ and $u_-$ will not. It follows clearly that, if a specific solution $y$ of (1) is approximated at $z \neq 0$ by a linear combination $c_1u_+ + c_2u_-$, then it cannot be approximated by the same linear combination at $z \exp(2\pi i)$. The concept of approximation here involved must be domain-dependent.

That is the basic Stokes Phenomenon.

The obvious next question is: why should we want to characterize an entire function by multivalued functions?

## 3. Natural Metric

There are, in fact, very strong mathematical and scientific reasons for such a choice. From the mathematical point of view, the functions $u_+$ and $u_-$ are exponentially large or small in $|z|$ (except when $x$ is purely imaginary), and that property of exponential "dominance" or "recessicn" offers a striking and powerful, qualitative characterization of the solution structure of (1). Similarly, $u_+$ and $u_-$ are oscillating functions of $z$ (except when $x$ is purely real) and that is also a salient mathematical property of the solutions of (1).

In science, the functions $y(z)$ in (1) arise usually as Fourier components, i.e., as the spatially varying factors of functions $y(z)\exp(-i\omega t)$ simple harmonic in time. Except when $x$ is purely real, the functions

$$u_+ e^{-i\omega t} = z^{-1/4} e^{x-i\omega t}$$

$$u_- e^{-i\omega t} = z^{-1/4} e^{-x-i\omega t} ,$$

represent waves, and so do the solution representations for which $(c_1u_+ + c_2u_-)\exp(-i\omega t)$ stand as symbols. The approximate wavelength $3\pi/z^{1/2}$ varies with spatial position, and so do the approximate amplitudes $c_1z^{-1/4}$ and $c_2z^{-1/4}$, as is usual in science. When $x$ is purely imaginary, those functions represent purely progressive waves and when both Re $x$ and Im $x$ are nonzero, they represent growing and damped waves, respectively; for real $x > 0$, e.g, the recessive function $u_-$ thus describes a wave shadow. This wave character of the solutions of (1) is their most important property, by far, in the scientific context and in many instances, it is their only scientifically relevant property. The representation by multivalued functions is the only way in which it can be displayed with great clarity.

These considerations carry over, first of all, to the general linear differential equation of second order,

$$\varepsilon^2 \, d^2w/dz^2 - p(z)\, w(z) = 0 \,, \tag{3}$$

with analytic coefficient function $p(z)$ and parameter $\varepsilon$. The corresponding wave approximations (often called WKB or Liouville-Green approximations) are

$$v_{\pm} = p^{-1/4} e^{\pm x}, \qquad x = \frac{1}{\varepsilon} \int^{z} [p(s)]^{1/2} ds \tag{4}$$

and since they are rigorous approximations [16], they are again first instalments towards, and symbols for, exact solutions of (1).

If $p(z)$ has a root $z_0$ ("turning point") in the complex z-plane and is analytic at $z_0$, then $w(z)$ is also analytic at $z_0$, but $v_+$, $v_-$ have a branch point there (unless the root has multiplicity $4^n$). Normally, therefore, they are then approximations to solutions of (3) only for $z \neq z_0$ and sufficiently large $|x-x(z_0)|$, and they approximate locally singlevalued solutions by locally multivalued functions $v_+$, $v_-$, which can therefore be only domain-dependent approximations. Hence, the Stokes Phenomenon of a multivalued representation of singlevalued functions arises again.

A very similar situation arises near any singular point $z_*$ of the differential equation (3). If $x(z_*)$ exists, (4) gives again approximations to solutions for sufficiently large $|x-x(z_*)|$ [16, pp. 222-224]. The solutions $w(z)$ themselves are then usually multivalued near $z_*$, but the multivaluedness of $v_+$ and $v_-$ differs from that of the solutions. The approximations by $v_+$, $v_-$ must therefore be domain-dependent and the Stokes Phenomenon arises.

But equally, the solution representations symbolized by (4) are precisely those displaying clearly the fundamental solution properties of wave character and dominance and recession. The Stokes Phenomenon is a necessary, and rather economical, price for the representations we need most of all.

Another conclusion must be that the variable $x$ in (4) (and of which that in (2) is but the simplest, special case) plays a fundamental role in the structure of wave equations of the form (3). It has therefore been called [23] the "natural metric" of (3). A physicist would say that $x$ is more natural than $z$, which measures distance in rather extraneous units like cm or inches, because $x$ measures distance in the intrinsic unit of the local wavelength of (3).

It is perhaps more illuminating to observe that the representations symbolized by (4) are closely related to the canonical representation of Hamiltonian oscillators in terms of action and angle [4], and an exact natural metric of (3) would correspond to Hamilton's angle. For periodic oscillators, an exact angle variable is defined, but it can be difficult to establish its precise relation to clock-time, because that amounts to solving the oscillator equation completely. A simple, first approximation to the angle is therefore valuable in practice. For modulated oscillators, a unique angle in the classical sense is not definable, and a simple approximate standby for it is an all the more valuable tool. For wave equations, similarly, an exact wave length exists only for strictly periodic solutions, and its precise determination is also tantamount to solving (3). The approximate natural metric $x$ is a first installment towards this and the author has not yet come across an instance where it has not been a tool adequate for analysis.

It should be recalled at this point that the action-angle concept is not restricted to linear oscillators, it plays an even more crucial role in nonlinear Hamiltonian structures. Its decisive role in the modulation of completely nonlinear oscillators, even though a classical angle is not then definable, has been demonstrated, e.g., in [9]. The Stokes Phenomenon arises equally in the nonlinear context, because a characterization of the

oscillatory structure is then needed even more for understanding and prediction. Nor is the action-angle concept restricted to oscillator equations (3) of second order, it applies to systems of such equations and to higher-order differential equations, as exemplified in the theory of hydrodynamic stability [2], for instance.

The key role of a natural metric in such structures is underlined with even more force in partial differential equations describing waves in several dimensions. A particularly fruitful approach to such waves is by Hamilton-Jacobi theory, which uses "ray" equations to construct a natural metric, usually called characteristic or phase or action function [5]. This complex phase is a multivalued function of position in space in a way quite analogous to the natural metric $x$ of (3) and serves again for a nonuniform and multivalued approximation and representation of the wave structure that is the most important feature of the solutions. Once it is constructed, moreover, this phase function can be used with utmost economy and elegance [5] to obtain qualitative and quantitative information on wave spectra. In this connection, the nonuniformity of approximation turns out to be no obstacle, and the multivaluedness [19] and basic Stokes Phenomenon play a central role. Significantly, again, Hamilton-Jacobi theory is not restricted to linear wave equations.

In sum, there are overwhelming reasons for the choice of multivalued representations of wave structures in terms of a natural metric. It emerges, moreover, that the simple linear differential equation (3) offers a felicitous example displaying the essentials of a structure of great generality. Clarity will therefore be served best by conducting the following discussion of the Stokes Phenomenon entirely within the framework of the simple differential equations (3).

In this connection, the natural metric $x$ defined in (4) plays a further useful role. The equation (3) is called the normal form of the general linear ordinary differential equation of second order. This term involves misleading associations, however, because such a differential equation has a large family of normal forms, linked by Liouville's transformation [16, p. 191]. In this way, for instance, Airy's equation (1) can be rewritten as a Bessel equation. In fact, the family of normal forms is so large that the relation between two normal forms of the same differential equation may be hard to recognize without the explicit Liouville transformation between them. Indeed, one normal form of a differential equation may have turning points or singular points that appear absent in another of its normal forms! This profusion of different normal forms contributes much to the difficulty of distinguishing basic features from more specialized ones in the literature.

It would be helpful, therefore, to have a "standard" form of the differential equation on which a fruitful discussion can be based from the start. Unfortunately, this depends on the questions at issue. When wave structure is of central concern, however, the natural metric offers a definite, standard form. In terms of

$$W(x) = [p(z)]^{1/4} w(z) ,$$

with again $x = \frac{1}{\varepsilon} \int^{z} [p(s)]^{1/2} ds$ , (3) becomes

$$\left.\begin{aligned} d^2W/dx^2 - (1+\varepsilon^2\psi)\, W &= 0 , \\ \psi(x) &= -p^{-3/4} d^2(p^{-1/4})/dz^2 . \end{aligned}\right\} \qquad (5)$$

This standard form of (3) already shows clearly that the differential equation is effectively a wave (or oscillator) equation where $\varepsilon^2|\psi(x)| << 1$ (and $x$ is not purely real), that the natural metric $x$ there measures distance in

approximate local wavelengths and that approximations of the form (4) are there plausible.

Furthermore, the superficial distinction in the literature between turning points of (3), i.e., roots of $p(z)$, and singular points of (3) disappears: both are singular points of the standard form (5) and the similarity in their significance for the solution structure needs no further explanation.

All of this shows that, when wave aspects are the central concern, as is usual in science, then the standard form (5) offers the best starting point for a direct road to an understanding of solution structure and the arbitrary form (3) is best forgotten for a simple explanation of the Stokes Phenomenon.

## 4. Stokes Lines

It is easier now to return to the domain-dependence of representations symbolized by the multivalued wave-approximations (4). It will be recalled from Section 2 that, if a solution $y(z)$ of (1) is approximated at a point $z \neq 0$ by a linear combination $c_1u_+ + c_2u_-$ of the wave approximations $u_+$, $u_-$ in (2), then its apprximation at $z \exp(2\pi i)$ as a linear combination of $u_+$ and $u_-$ cannot have the same coefficients $c_1, c_2$ because $z = 0$ is an ordinary point of (1), but a branch point of $u_+$ and $u_-$.

To learn more about such change in the coefficients, it is best to start from the natural metric and to focus attention first on the case of a differential equation (5) with only one, finite singular point, which can be taken at $x = 0$. Let $r(\varepsilon x)$ denote a definite branch, defined on the complex plane of $x$ cut from $0$ to $\infty$, of the fourth root of the coefficient function $p(z)$ from (3) and let $\psi(x)$ in (5) be understood as the corresponding branch. Then the wave approximations for (5) corresponding to

$v_+$, $v_-$ in (4) are

$$V_+(x) = e^x , \qquad V_-(x) = e^{-x} . \tag{6}$$

More precisely, the Liouville-Green or WKB theorem [16, pp. 222-224] states that (5) has a fundamental system of solutions

$$W_+ = a(x;\varepsilon)e^x , \qquad W_- = b(x;\varepsilon)e^{-x} \tag{7}$$

with the property that $|a|$ and $|b|$ are bounded for large $|x|$. For a very large class of functions $p(z)$ in (3), moreover, $a(x;\varepsilon)$ and $b(x;\varepsilon)$ tend to limits as $|x| \to \infty$ for fixed $\varepsilon$ with arg $x$ an integer multiple of $\pi$ [16, pp. 222-224].

For large $|x|$, therefore, any one solution $W(x)$ of (5) is approximated by a linear combination $c_1V_+ + c_2V_-$ in which $c_1$ and $c_2$ are such limits of $a(x;\varepsilon)$ and $b(x;\varepsilon)$, respectively. The situation of Section 2 is now reversed: the approximating functions $V_+$ and $V_-$ are entire, but $\psi$ has a branch point at $x = 0$ and hence, $W(x)$ has normally also a branch point there. Therefore, the coefficients in the approximation $c_1V_+ + c_2V_-$ to $W$ must jump across the cut. That is the Stokes Phenomenon.

In a general way, the choice of a cut is rather arbitrary, but in the context of approximation, there is a natural preference, which Stokes already referred to [21]. For $V_-(x)$, it is the positive real axis of $x$ because that is the line of deepest recession of $V$. Indeed, if $W \sim c_1V_+ + c_2 V_-$ and $c_1$ is not exactly zero, then in domains where $V_+$ is dominant, $c_2V_-$ is 'invisible' in what has come to be called Poincaré's sense of approximation [16] (and has often been relied on far beyond his intention). 'Invisibility' means here that, however the first approximation be refined to successively higher approximations in the technical sense just referred to, $c_2V_-$ will always remain submerged in the approximation error. Similarly, the negative real axis is the preferred cut for $V_+$. (There is no satisfactory way in

which the same cut can be used for both and therefore, no single branch of $r(\varepsilon x)$ in the way first defined is adequate, but the complication can be kept minimal by use of a definite branch of $r(\varepsilon x)$ defined [12] on a Riemann surface sector with winding point at $x = 0$ and opening angle $3\pi$.)

If the rationale just sketched could be relied on, the cut for $V_-$ could be chosen along any line in the domain $|\arg x| < \pi/2$ of recession of $V_-$ because the technical 'invisibility' of $V_-$ stems wholly from the contrast between dominance and recession. In the literal sense of asymptotic expansions often pushed to excess in the mathematical literature, this leads to the theorem "for arbitrarily small $\delta > 0$, and sufficiently large $|x|$, $W \sim c_1 V_+$ for $|\arg x| \leq (\pi/2)-\delta$, unless $c_1 = 0$ exactly, and higher approximations have the same sector of validity." In short, the change in $c_2$ appears only suddenly as the imaginary x-axis is approached on which $V_+$ and $V_-$ are "balanced" because $|V_+| = |V_-| = 1$ there.

This is totally unrealistic, however. The admission can no longer be postponed that this account has sinned by discussing approximation without reference to the meat of that subject, which is [16] the bound on the approximation error. (The reason is that realistic error bounds [16] demand an excessive volume of hard analysis for a "simple explanation.") If the theorem just stated is applied with small $\delta$, the error bound [15] comes to depend strongly on $\delta$ and for small $\delta$, $W \sim c_1 V_+$ is even a rough approximation only for values of $|x|$ too large to be of any interest. An instructive example [15] of this is quoted in the Appendix. For any sensible purpose, the value of $\delta$ in the theorem may need to be very substantial [15], and this explains better why the positive real axis should be the preferred cut in regard to $V_-$, and the other real half-axis, in regard to $V_+$. The images of these half axes in the Liouville family of z-planes are the original

Stokes lines [16, p. 518], but in recent times, the name Stokes line has come to be more commonly attached to the images of the imaginary half-axes, where the solutions W are "balanced," because the latter lines need to be referred to more often in physics and in connection theory and appear as the characteristic lines (or "rays") of Hamilton-Jacobi theory. (Of course, in cases where an exact natural metric exists, it would be even better to choose the cuts on, and attach these names to, the images of the half-axes of the exact natural metric.)

Historically, the Stokes Phenomenon arose first in the analysis of the classical functions of mathematical physics, which possess exact and concrete integral representations in terms of elementary analytic functions. Application of the method of steepest descent to those exact integral representations then led automatically to asymptotic approximations of wave character. They were seen immediately to explain the qualitative and quantitative structure of the classical functions over most of their domain. Everybody knows, for instance, that for Bessel functions of low order and real argument, the interval on which their oscillatory character dominates is huge by comparison with the interval in which that character is not apparent. Both rigorously and intuitively, the wave-type approximations to the classical functions were seen in this way to be such an obvious necessity that the question at the end of Section 2 (and hence, all of Section 3) was redundant. On the other hand, the Stokes Phenomenon arose in that way as an erudite, technical byproduct of the steepest-descent approximation of integral representations explainable only in terms of complicated analysis. This impression of a large technical apparatus is enhanced by the fact that whole asymptotic expansions can be, and are usually, written out in elaborate formulae for the special functions. Stokes [21] observed that the Phenomenon

is a general concomitant of a class of technical operations, which arise often. All this made access to the Phenomenon tortuous in the literature.

Apart from the omission of proofs and error bounds, this simpler explanation has also relied on the assumption that the differential equation (5) has only one finite singular point. If there are more, any description of approximation must be more elaborate [18]. This is aggravated further when the differential equation [9] depends on a parameter at a value of which two [12], [17] or more [22] singular points coalesce, a situation of considerable scientific interest, but beyond the present scope. (It can also happen [14] that two singluar points are close to each other without coalescence.)

It should be remarked that the Stokes Phenomenon occurs also in other contexts. J.W. Gibbs considered [3] its relevance to the representation of real-valued, but not real-analytic, functions -- such as a periodic square pulse -- by Fourier series. In reality, the series must be truncated to approximate the square pulse by a partial sum, which is an entire analytic function. A square pulse, on the other hand, is the imaginary part of a multivalued analytic function and in this light, the Gibbs Phenomenon for the square pulse is a symptom of the Stokes Phenomenon.

A different analog of Stokes' Phenomenon can be exemplified at the hand of the Hankel functions of order zero as solutions of Bessel's equation of that order. A good definition of those functions is in terms of an integral representation [16, pp. 239-240]. It shows the Hankel functions to have a branch point at $z = 0$ and hence, their understanding requires a Riemann surface. Each sheet, however, is a replica of the complex plane and so the same definition of the Hankel functions is used on each sheet. The integral representation of $H_0^{(1)}(z)$ on sheet $m$ is then <u>not</u> the analytic continuation on the Riemann surface to sheet $m$ of the integral representation of

$H_0^{(1)}(z)$ on sheet $k \neq m$. On each sheet, on the other hand, the two Hankel functions -- as defined on that sheet -- are a fundamental system of Bessel's equation. Hence, there must be "connexion formulae" of the form

$$H_0^{(1)}(ze^{2m\pi i}) = c_m H_0^{(1)}(z) + d_m H_0^{(2)}(z)$$

[16, pp. 239-240]. If a definite Bessel function is represented as a linear combination of Hankel functions, it follows that the coefficients must change from sheet to sheet. The analogy with the Stokes Phenomenon is close because the integral representations on each sheet are chosen to obtain there the simplest, clearcut properties of dominance and recession for the Hankel functions. In other words, they are chosen to agree with the wave representations sympbolized by $V_+$ and $V_-$. That must cause a Stokes Phenomenon and connexion formulae are often interpretable in this light. More precisely, they look, at this point, like uniform extensions of the asymptotic Stokes Phenomenon by means of exact integral representations. The apparent nonuniformity of the Stokes Phenomenon, however, is caused only by its still incomplete discussion.

## 6. Connection

The question remains how the changes in the coefficients of the wave-approximation $c_1V_+ + c_2V_-$ to a solution $W$ of the wave equations (5) are to be computed. It has a far greater scientific importance than may appear at first sight. The solution $W$ is normally hard to observe, in fact, it is unobservable by axiom in quantum mechanics, but even in sciences based on classical physics, it can rarely be observed with any precision at all, except in laboratory experiments specially designed for that purpose. But, $c_1$ and $c_2$ play a more serious role. The imaginary axis of $x$ is distinguished by the fact that $|V_+| = |V_-| = 1$ so that $V_+$ and $V_-$ represent purely

progressive waves on it, in terms of which radiation and reflection conditions are most simply and directly interpreted. The coefficients $c_k$ are the complex amplitudes of those progressive waves, i.e., $|c_k|$ are the real amplitudes and $\arg c_k$ are the phase constants. An abiility to relate the respective $(c_1,c_2)$ - pairs in different parts of the domain of the wave equation therefore translates immediately into a prediction of the scattering matrix telling us the reflected wave amplitude and phase caused by a given incident wave amplitude and phase, or vice versa. Those we have a much better hope of observing, they are axiomatic observables even in quantum mechanics and are realistically desirable predictions in many sciences. In turn, moreover, the scattering matrices determine the spectrum, which determines many further realistic predictions, and aspects of which are normally among the prime observables. It should be emphasized here that scattering matrices and spectra depend, not on the complex amplitudes $c_k$, but on the relation between them, that is, on the quantitative resolution of the Stokes Phenomena. The amplitudes themselves are normalization constants and, like the solutions $W$ of the wave equation, are normally redundant in the scientific context.

Establishment of the relation between respective amplitude pairs $(c_1,c_2)$ in different parts of the domain has come to be known as "connection." The first mathematically reliable method for it is due to Zwaan [24], but its range of application is too restricted. A more general method was pioneered by Langer [7] and has dominated 'mathematical asymptotics' for half a century on account of its triumph of uniform approximation of the solutions. This gives illuminating information, for instance, it shows just how the transition from light to shadow looks in detail, whereas the differential equation itself tells only where it happens. It is also essential for the quantitative

mastery of special functions [16]. However, it offers only an indirect and cumbersome approach to connection and hence, also to scattering, to spectra and to other aspects of more primary scientific significance. Still, this cumbersome approach has been brought to high perfection [18] and has even been generalized to two and three dimensions [6, 8, 20], albeit without proof because it gets prohibitive unless it starts from the assumption that the local variation of the solution is effectively only in one of the space dimensions. It is also committed to the limitations of the literal Poincaré sense of approximation with its potential lack of any realism (Appendix) and has so far failed where science needs to know both balanced (or dominant) and also recessive functionals of a wave equation [10, 11]. In sum, it has done marvelous things, but not for the explanation of the Stokes Phenomenon, which emerges from it only as a very late, technical byproduct of a large volume of hard analysis.

In recent years, however, a simpler and more direct approach to connection has been found which clarifies the Stokes Phenomenon further. Its simplest statement [16, p. 481] arises for an ordinary or isolated singular point of $p[z]$ in (3), which may be placed at $z = 0$, for brevity. For regular singular points, in particular, Frobenius' method [16, p. 149] shows that the differential equation then leads simply and quickly to an "indicial equation" the roots $\alpha_1, \alpha_2$ of which are the exponents of a fundamental system

$$w_1(z) = z^{\alpha_1} F_1(z), \qquad w_2(z) = z^{\alpha_2} F_2(z)$$

with entire functions $F_i(z)$ (provided $\alpha_1 - \alpha_2$ is not an integer or zero). Obviously, those constant exponents $\alpha_1, \alpha_2$ describe the multivaluedness of the $w_i$ completely globally. Each of the WKB solutions $W_+$, $W_-$ must be a linear combination of $w_1$, $w_2$ whence the Stokes relation

between $a(\infty;\varepsilon)$, $b(\infty;\varepsilon)$ and $a(\infty\, e^{2\pi i};\varepsilon)$, $b(\infty\, e^{2\pi i};\varepsilon)$ is deducible immediately [16, p. 481].

This sounds almost too good to be true and so it is, in a sense [16, pp. 480, 482]. It mounts to putting Zwaan's method [24] in a nutshell, with extension to isolated singular points of $p(z)$ in (3). It relies basically on the following fact. The circuit exponent of a multivalued function $W(x)$ is defined by tracing a circuit once anticlockwise around a branch point and letting $W_1(x)$ denote the value of $W$ at the start $x$ of the tracing and $W_2(x)$, that found at the end; then

$$C = \log (W_2/W_1)$$

is the circuit exponent. For an ordinary point or isolated singular point of the differential equation, the circuit exponent $C$ of a solution is clearly an invariant independent of $x$. Admittedly, access to this invariant is known only for ordinary and regular singular points [16, p. 480, 482]. For other singular points, for instance, for logarithmic turning points of (3), the circuit exponent is <u>not</u> invariant and the simple argument collapses [16, p. 480, 482]; such cases had not been found tractable by Langer's method either.

Multivaluedness, however, must be a global structure of inherent invariance, even if no simple way be known of defining the invariant quantitatively. An extension of Olver's idea has been found [13] which applies to a very general class of even highly irregular singular points (and turning points) of wave equations. When such a point is adequately separated from any others in the plane of the natural metric, it suffices to define the branch $r(\varepsilon x)$ of $[p(z)]^{1/4}$ characterizing the wave equation (5) <u>purely locally</u> in an arbitrarily small neighborhood of that singular point, placed at $x = 0$, say. The singularity there can be left virtually arbitrary,

certainly, the circuit exponent can vary with $|x|$, powers and logarithms are merely the most elementary special examples of the wide variety of functions admitted. Only three conditions, in fact, are used in [13]: First, $r(\varepsilon x)$ is analytic on the cut neighborhood (or more precisely, sectorial Riemann surface element) where it is defined. Secondly the limit

$$\lim_{|x| \to 0} \frac{x}{r} \frac{dr}{dx} = \gamma \in \mathbb{C} \tag{8}$$

of a circuit exponent at the singular point itself exists. And finally, $x = 0$ is the image of a finite turning or singular pont of (3) (which makes $\mathrm{Re}\ \gamma < 1/2$ [11]). Those strictly local properties are shown to determine the asymptotic relations between the wave amplitude functions $a(x;\varepsilon)$, $b(x;\varepsilon)$ in (7) far away from the singular point, as $|x| \to \infty$ with $|\varepsilon x| \to 0$ and $\arg x$ an appropriate multiple of $\pi$.

This is as much information as is usually justifiable, because the wave equation itself is only an imperfect model of reality. Whether more than a first approximation to a prediction of such a model can be relevant, can be decided only in the light of further examination of the quality of the model. This underlines the importance of the first approximation to connection, which is the relation between the values for different integers $m$ of the limits

$$c_1 = a(\infty e^{2m\pi i};\varepsilon), \qquad c_2 = b(\infty e^{(2m-1)\pi i};\varepsilon)$$

in (7). Those relations turn out [13] to depend only on the limit $\gamma$ in (8).

In short, even in a very general context, the asymptotic connection which quantifies the Stokes Phenomenon can be read directly off the most local properties at the singular point itself of the wave equation. The Phenomenon is anchored in the multivaluedness structure inherent in singular points of wave equations (5) in the natural metric. This is a global aspect of structure not tied to any notion of approximation. It is a matter of

tradition, more than substance, that we attach Stokes' name to an asymptotic aspect of that structure.

## Appendix. Olver's Example

A simple example [15] illustrating the disasters that can arise from reliance on the common notion that recessive terms are very small, deserves wider dissemination. Repeated integration by parts shows easily that the function

$$I(n^2) = \int_0^{\pi} \frac{\cos(nt)}{t^2+1} dt$$

has for large integer $n > 0$ the asymptotic expansion

$$I(n^2) \sim (-1)^{n-1} \sum_{k=1}^{\infty} \lambda_k n^{-2k} \qquad (A1)$$

with recursively defined coefficients $\lambda_k$ of which the first three are [15]

$$\lambda_1 = 0.05318, \qquad \lambda_2 = 0.04791, \qquad \lambda_3 = 0.0895.$$

Since those coefficients are of similar size, the expansion goes effectively just in steps of $n^{-2}$, which makes for a gratifying decrease in the contributions of successive terms as soon as $n^2$ is large. For $n^2 = 100$, e.g., the size of each term is less than 0.02 times that of the preceding term, and the quantitative efficiency of the expansion is excellent. To seven decimals, (A1) gives

$$I(100) = - 0.000\ 527\ 1\ , \qquad (A2)$$

where the third term in (A1) contributes only + 1 in the seventh decimal and the fourth term has no influence at all.

The actual value of I(100) to seven decimals is [15]

$$I(100) = - 0.000\ 455\ 8\ . \qquad (A3)$$

The error in (A2) is 0.000 071 3, or 15.6%, and is larger than all but the first approximation predicted by (A1)! Observe that, as is the norm in approximation theory, (A1) is rigorous and gives no hint at all of its failure.

The reason for the large error in (A2) is that $I(n^2)$ contains also a term $(\pi/2)\exp(-n)$, which is highly recessive at $n = 10$, but has there just the value 0.000 071 3 of the error in (A2) to seven decimals.

Rigor is tied to definitions, which may be inappropriate.

Acknowledgements. The author is indebted to Profs. T.B. Benjamin and J. Bona for helpful remarks.

## References


[1] C. Chester, B. Friedman and F. Ursell, An extension of the method of steepest descents, Proc. Camb. Phil. Soc., 53(1957), pp. 599-611.

[2] P.G. Drazin and W.H. Reid, Hydrodynamic Stability, Cambridge Univ. Press, London, 1981.

[3] J.W. Gibbs, Fourier's series, Nature, 59(1898), p. 200 and 59(1899), p. 606; also The Collected Works, Vol. II Pt. II, Longmans, New York, 1928, pp. 258-260.

[4] H. Goldstein, Classical Mechanics, Addison-Wesley, Cambridge, 1950.

[5] J.B. Keller, Semiclassical mechanics, SIAM Rev. 27(1985), pp. 485-504.

[6] Yu. A. Kravtsov, A modification of the geometrical optics method, Izv. Vyss. Ucbn. Zaved. Radiofizika, 7(1964), pp. 664-673.

[7] R.E. Langer, On the asymptotic solutions of ordinary differential equations, with an application to the Bessel functions of large order, Trans. Amer. Math. Soc., 33(1931), pp.23-64.

[8] D. Ludwig, Uniform asymptotic expansions at a caustic, Comm. Pure Appl. Math., 19(1966), pp. 215-250.

[9] R.E. Meyer, Adiabatic Variation, Part V, Nonlinear near-periodic oscillator, ZAMP 27(1976), pp. 181-195.

[10] R.E. Meyer, Exponential asymptotics, SIAM Rev., 22(1980), pp. 213-224.

[11] R.E. Meyer, Wave reflection and quasiresonance, Theory and Applications of Singular Perturbations, W. Eckhaus and E.M. de Jager, eds., Lecture Notes in Mathematics, Vol. 942, Springer-Verlag, Heidelberg, 1982, pp. 84-112.

[12] R.E. Meyer and J.F. Painter, Irregular points of modulation, Adv. Appl. Math. 4 (1983), pp. 145-174.

[13] R.E. Meyer and J.F. Painter, Connection for wave modulation, SIAM J. Math. Anal. 14, (1983), pp. 450-462.

[14] R.E. Meyer, Quasiresonance of long life, J. Math. Phys. 27(1986), pp. 238-248.

[15] F.W.J. Olver, Error bounds for asymptotic expansion, with an application to cylinder functions of large argument, Asymptotic Solutions of Differential Equations and Their Applications, C.H. Wilcox, ed., John Wiley, New York, 1964, pp. 163-183.

[16] F.W.J. Olver, Asymptotics and Special Functions, Academic Press, New York, 1974.

[17] F.W.J. Olver, Second-order linear differential equations with two turning points, Philos. Trans. Roy. Soc. London Ser. A, 278(1975), pp. 137-174.

[18] F.W.J. Olver, General connection formulae for Liouville-Green approximations in the complex plane, Philos. Trans. Roy. Soc. London Ser. A, 289 (1978), pp. 501-548.

[19] I.C. Percival, Semiclassical theory of bound states, Adv. Chem. Phys., 36(1977), pp. 1-61.

[20] M.C. Shen and J.B. Keller, Uniform ray theory of surface, internal and acoustic wave propagation in a rotating ocean or atmosphere, SIAM J. Appl. Math., 28(1975), pp. 857-875.

[21] G.G. Stokes, On the numerical calculation of a class of definite integrals and infinite series, Trans. Camb. Philos. Soc. 9(1850) pp. 166-187, reprinted in Mathematical and Physical Papers, Vol. 2, Camb. Univ. Press, London, 1883, pp. 329-357.

[22] F. Ursell, Integrals with a large parameter. Several nearly coincident saddle points, Proc. Camb. Phil. Soc., 72(1972), pp. 49-65.

[23] A-M. Wazwaz and F.B. Hanson, Singular boundary resonance with turning point resonance, SIAM J. Appl. Math 46(1986), pp. 962-977.

[24] A. Zwaan, Itensitaeten in Ca-funkenspektrum, Arch. Neerlandaises Sci. Exactes Natur. Ser. 3A, 12, (1929), pp. 1-76.

END

DATE

FILMED

DTIC

JULY 88